\*Note: I’ve noticed in my Excel sheet that sometimes my graphs next to the pivot charts don’t show up right away – they are there, but sometimes I have to zoom in or out for the graph to actually appear.

One conclusion we can draw from this data is that a large majority of the campaigns included in this dataset are based in the US (763 out of 1,000). We can also see that, overall, the category with the most crowdfunding campaigns is theater (344 out of 1,000). Finally, we can see that campaigns created in the month of July had the highest number of successful campaigns compared to campaigns created in other months, but it also had the highest overall number of campaigns created compared to other months.

One limitation of this dataset is that there is limited information on who was running each campaign and how many resources they started with. Campaigns run by existing companies or corporations will be in a better starting position than a campaign run by an individual(s), and will likely contribute to the campaign’s likelihood of success.

Additionally, the data does not tell us how any of the funds were used for each campaign. The data can tell us how much each campaign received in funds, how many backers there were, what category/subcategory it falls under, and the outcome of the campaign, but it does not provide any real information about how the campaign got to its outcome. Knowing how a campaign uses its funds is arguably the most important piece of information when determining what contributed to a campaign’s success or failure.

Instead of outcome based on goal, we could create a table and line or bar graph of outcome based on pledged. This could be set up in the same way that outcome based on goal is set up, but replacing “goal” with “pledged”. This would give us a better idea of how a campaign performs based on the amount of money actually raised rather than the amount of money they wanted to raise. Depicting the outcome based on goal is not as useful as depicting outcome based on pledged because the goal amount is typically not an accurate depiction of the funds available to each campaign.

We could also create a table and line graph showing the outcome based on percent funded. We could create a table similar to the outcomes based on goal table, but instead of the “Goal” column, we have a “Percent Funded” column with 10 percentage point increments (0-9%, 10-19%, 20-29%, …, greater than 100%). This would show if there were any correlation between a campaign’s outcome and how much of their goal was met. This could also be depicted as a bar graph instead of a line graph, with the percent funded on the X-axis, the percentage of the outcome on the Y-axis, and different colored bars depicting each outcome (successful, failed, cancelled).

**Statistical Analysis**

For this dataset, the median summarizes the data better than the mean. There are too many potential outliers for the mean to represent the data realistically, and since the median is much less affected by outliers, it provides a better depiction of this dataset.

There is more variability in the number of backers for successful campaigns than unsuccessful campaigns. We know this because the variance of the number of backers and the standard deviation is larger in successful campaigns than unsuccessful ones. This makes sense because there is more of a spread in the data for successful campaigns than unsuccessful ones (there is a greater difference between the min and max number of backers for successful than unsuccessful campaigns).